PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-258768
(43)Date of publication of application : 03.10.1997

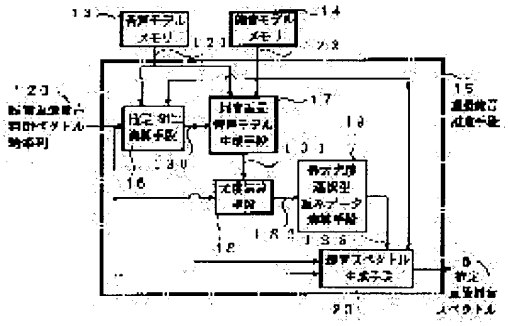
(51)Int.CI. G10L 3/00
G10L 3/02

(21)Application number : 08-068210
(22)Date of filing : 25.03.1996

(71)Applicant : MITSUBISHI ELECTRIC CORP
(72)Inventor : SUZUKI TADASHI

(54) UNDER-NOISE VOICE RECOGNIZING DEVICE AND UNDER-NOISE VOICE RECOGNIZING METHOD

(57)Abstract:
PROBLEM TO BE SOLVED: To suppress the degradation of recognizing performance caused by the fluctuation of environmental noise and the fluctuation of distance between noise source and a voice input microphone by obtaining estimated superposed noise spectrum using a noise model and a noiseless voice model.
SOLUTION: An estimated S/N computing means 16 computes an estimated S/N 130 to each vector of a noise superposed voice feature vector time series 120 using a noiseless voice model 122 stored in a memory 13 and a noise model 121 stored in a memory 14. A noise superposed voice model generating means 17 synthesizes the noiseless voice model 122 and the noise model 121 according to the estimated S/N 130 to generate a noise superposed voice model 131. A noise spectrum generating means 20 generates estimated superposed noise spectrum 5 using the noise superposed voice feature vector time series 120, the estimated S/N 130, weight data 133 which is the output of a maximum likelihood selecting type weight data computing means 19, and the noise model 121, and outputs it to a noise spectrum eliminating means.



LEGAL STATUS

[Date of request for examination]	07.02.2000
[Date of sending the examiner's decision of rejection]	15.04.2003
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3452443
[Date of registration]	18.07.2003
[Number of appeal against examiner's decision of rejection]	2003-08495
[Date of requesting appeal against examiner's decision of rejection]	15.05.2003
[Date of extinction of right]	

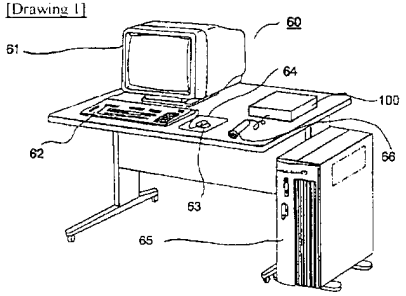
* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

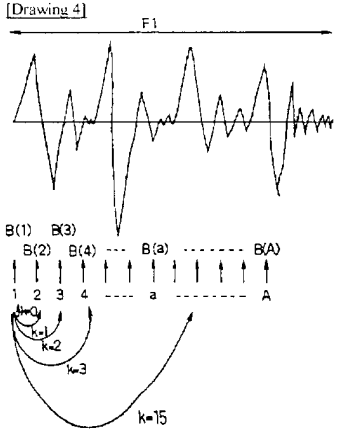
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

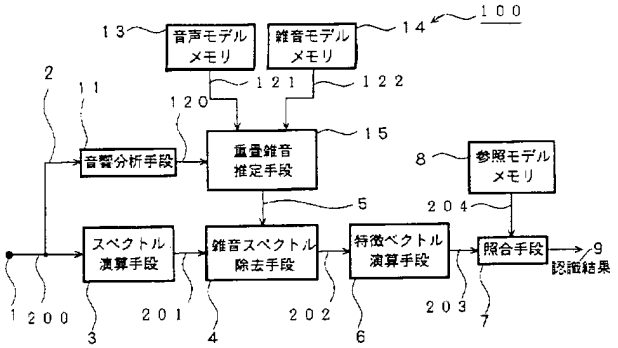
[Drawing 1]



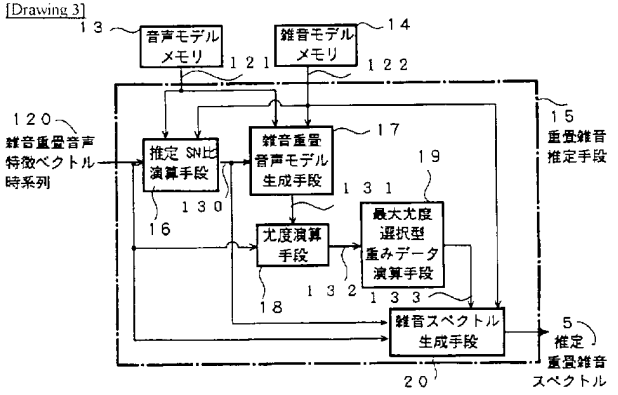
[Drawing 4]



[Drawing 2]



[Drawing 3]



[Drawing 6]

番号	種類	正規化相関係数ベクトル	正規化パワースペクトル
1	ホワイト雑音	V1	W1
2	ピンク雑音	V2	W2
3	工場雑音	V3	W3
⋮	⋮	⋮	⋮
n	電車雑音	Vn	Wn
⋮	⋮	⋮	⋮
N	オフィス雑音	VN	WN

[Drawing 7]

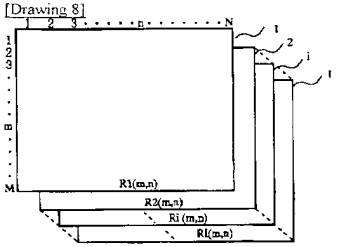
フレーム番号	分析フレーム	正規化自己相関係数ベクトル
1	F1	X1
2	F2	X2
3	F3	X3
⋮	⋮	⋮
i	Fi	Xi
⋮	⋮	⋮
I	FI	XI

[Drawing 5]

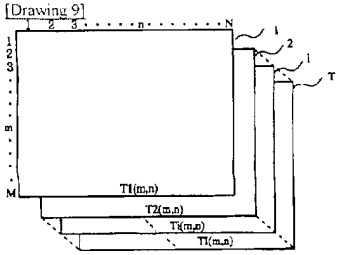
番号	種類	正規化相関係数ベクトル
1	母音a	S1
2	母音i	S2
3	母音u	S3
4	母音e	S4
5	母音o	S5
6	母音s	S6
⋮	⋮	⋮
m	子音f	Sm
⋮	⋮	⋮
M	子音z	SM

13

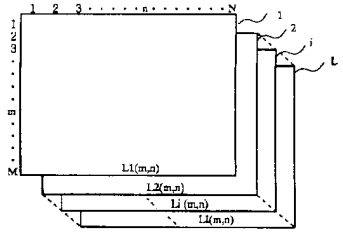
[Drawing 8]



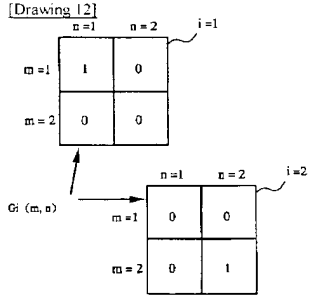
[Drawing 9]



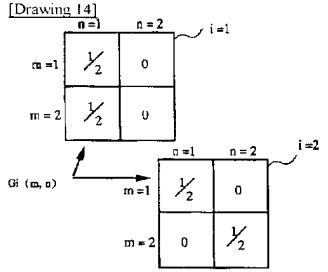
[Drawing 10]



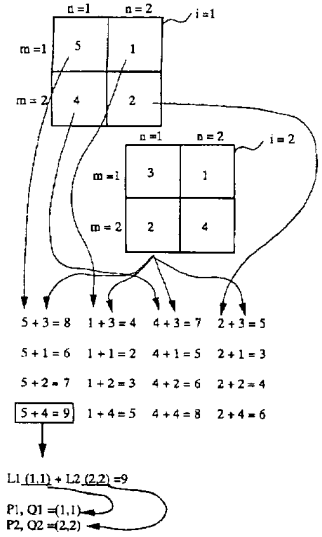
[Drawing 12]



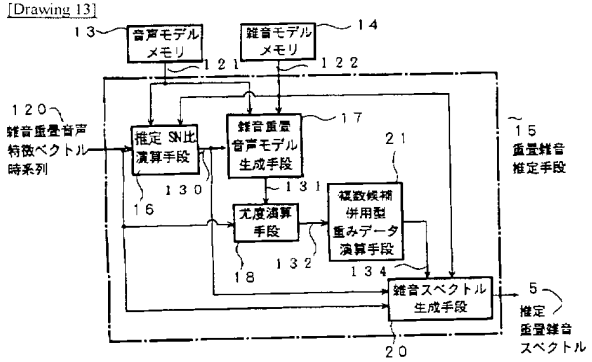
[Drawing 14]



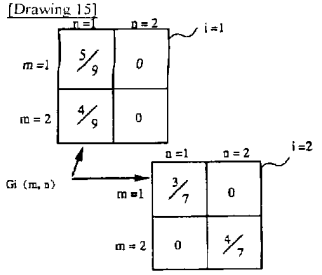
[Drawing 11]



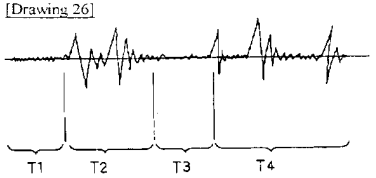
[Drawing 13]



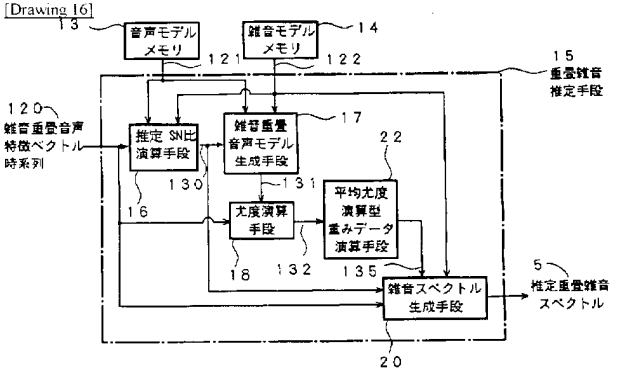
[Drawing 15]



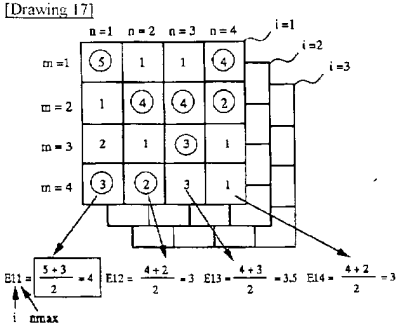
[Drawing 26]



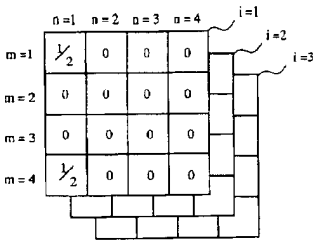
[Drawing 16]



[Drawing 17]



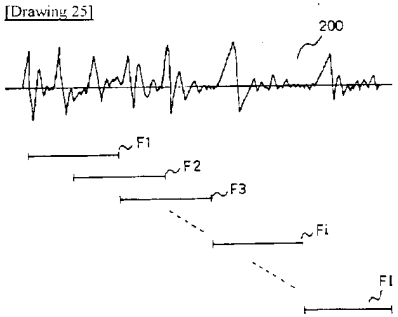
[Drawing 18]



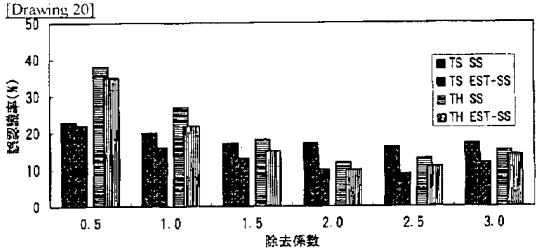
[Drawing 19]

音声データ	発声者：成人男性2名 話　者：音韻バランス492単語1セット（学習用） 100地名音声1セット（認識評価用）
付加雑音	PINK 雑音、工場騒音
音響分析	A/D：16bits 10kHz 高域強調：$1-0.95\cdot z^{-1}$ 分析窓／長期：Hamming 25.6msec／10.0msec LPC 分析：12次 FFT：256点
特徴ベクトル	LPC ケプストラム　1-16次 ＋デルタケプストラム　1-16次
音素片　HMM	連続分布、継続時間制御付き

[Drawing 25]

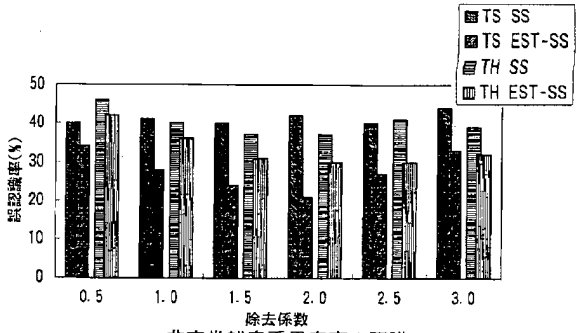


[Drawing 20]



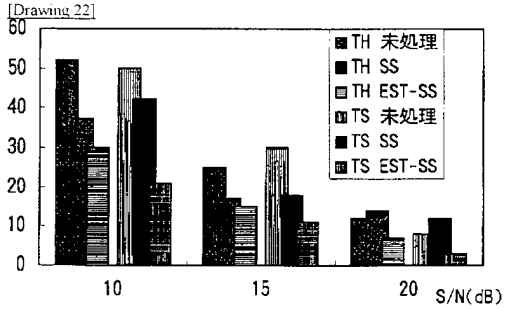
PINK雑音重畳音声の認識

[Drawing 21]



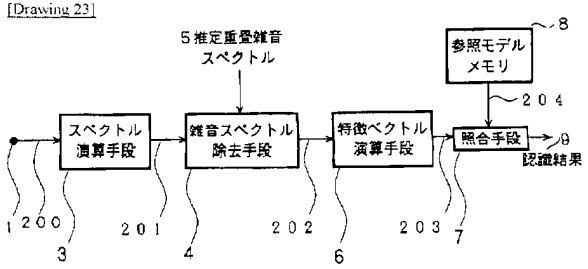
非定常雑音重畳音声の認識

[Drawing 22]

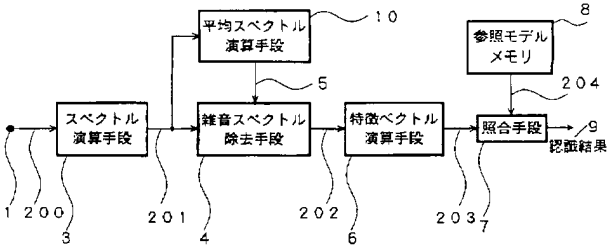


S/Nによる認識性能の変化

[Drawing 23]



[Drawing 24]



[Translation done.]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is drawing showing the speech recognition structure of a system in the gestalt 1 of implementation of this invention.
[Drawing 2] It is the block diagram showing the configuration of the gestalt of 1 operation of the bottom voice recognition unit of the noise in the gestalt 1 of implementation of this invention.
[Drawing 3] It is the block diagram showing the configuration of the gestalt of 1 operation of the superposition noise presumption means in the gestalt 1 of implementation of this invention.
[Drawing 4] It is drawing explaining the auto correlation coefficient in the gestalt 1 of implementation of this invention.
[Drawing 5] It is drawing showing the voice model memory in the gestalt 1 of implementation of this invention.
[Drawing 6] It is drawing showing the configuration of the noise model memory in the gestalt 1 of implementation of this invention.
[Drawing 7] It is drawing showing the normalization autocorrelation charge numerical vector of each analysis frame in the gestalt 1 of implementation of this invention.
[Drawing 8] It is drawing showing the presumed SN ratio in the gestalt 1 of implementation of this invention.
[Drawing 9] It is drawing showing the noise superposition voice model in the gestalt 1 of implementation of this invention.
[Drawing 10] It is drawing showing the likelihood in the gestalt 1 of implementation of this invention.
[Drawing 11] It is drawing showing how to ask for the sequence data in the gestalt 1 of implementation of this invention.
[Drawing 12] It is drawing showing the weight data in the gestalt 1 of implementation of this invention.
[Drawing 13] It is the block diagram showing the configuration of the gestalt of 1 operation of the superposition noise presumption means in the gestalt 2 of implementation of this invention.
[Drawing 14] It is drawing showing the weight data in the gestalt 2 of implementation of this invention.
[Drawing 15] It is drawing showing the weight data in the gestalt 2 of implementation of this invention.
[Drawing 16] It is the block diagram showing the configuration of the gestalt of 1 operation of the superposition noise presumption means in the gestalt 3 of implementation of this invention.
[Drawing 17] It is drawing showing how to calculate the maximum of the average likelihood in the gestalt 3 of implementation of this invention.
[Drawing 18] It is drawing showing the weight data in the gestalt 3 of implementation of this invention.
[Drawing 19] It is drawing showing the experiment conditions of this invention.
[Drawing 20] It is drawing showing the recognition result of the PINK noise superposition voice of this invention.
[Drawing 21] It is drawing showing the recognition result of the unsteady noise superposition voice of this invention.
[Drawing 22] It is drawing showing change of the recognition engine performance by the presumed SN ratio of this invention.
[Drawing 23] It is the block diagram showing the configuration of the gestalt of 1 operation of the bottom voice recognition unit of the noise using the conventional spectrum total method.
[Drawing 24] It is the block diagram showing the configuration of the gestalt of 1 operation of the conventional bottom voice recognition unit of the noise.
[Drawing 25] It is drawing showing the relation between the inputted noise superposition sound signal and an analysis frame.
[Drawing 26] It is drawing showing the relation between the voice section and the noise section.
[Description of Notations]
1 Input Edge, 3 Spectrum Operation Means, 4 Noise Spectrum Removal Means, 5 A presumed superposition noise spectrum, 6 A feature-vector operation means, 7 Collating means, 8 Reference model memory, 9 A recognition result, 10 Average spectrum operation means, 11 A sonagraphy means, 13 Voice model memory, 14 Noise model memory, 15 A superposition noise presumption means, 16 A presumed SN ratio operation means, 17 Noise superposition voice model generation means, 18 A likelihood operation means, 19 A maximum likelihood selection mold weight data operation means, 20 A noise spectrum generation means, 21 Two or more candidate concomitant use mold weight data operation means, 22 Average likelihood operation mold weight data operation means.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention is uttered under a noise environment and relates to the voice recognition unit and the speech recognition approach for the voice which the noise superimposed.
[0002]
[Description of the Prior Art] In speech recognition, collating processing based on the spectrum information on audio is performed. For this reason, the candidate for recognition may be carried out [voice / which was uttered in the suitable environment used for study of the voice model for collating, i.e., a different noise environment from the suitable environment where the voice data for collating was uttered,]. In this case, the recognition engine performance deteriorates greatly under the effect of the ambient noise signal superimposed on a sound signal.
[0003] There is the technique of removing the component of this noise on a power spectrum, assuming that the noise same also during the voice section is overlapped, after learning the aspect of a noise from the non-voice section in an input signal to this problem. Drawing 23 is an example of the block diagram of the bottom voice recognition unit of the noise using the spectrum total method shown in reference "voice acoustical engineering lecture ** revised [edited by Acoustical Society of Japan]" (Kazuo Nakada, Corona Publishing, P.130-131). A spectrum operation means for 3 to perform a analysis of a spectrum in drawing to the noise superposition sound signal 200 inputted from the input edge 1, and to output the noise superposition voice spectrum time series 201, 4 removes a noise component from the noise superposition voice spectrum time series 201 using the presumed superposition noise spectrum 5 prepared beforehand. A noise spectrum removal means to output the noise rejection voice spectrum time series 202, A feature-vector operation means by which 6 searches for the noise rejection voice feature-vector time series 203 from the noise rejection voice spectrum time series 202, 7 is a collating means to output the category of the voice model 204 for collating with which collating processing with the voice model 204 for collating memorized by the noise rejection voice feature-vector time series 203 and the reference model memory 8 is performed, and likelihood becomes high most as a recognition result 9.
[0004] Moreover, drawing 24 is an example of the block diagram of the bottom voice recognition unit of the noise using the average spectrum operation means 10 as a means to generate the presumed superposition noise spectrum 5, in the voice recognition unit of drawing 23 . The average spectrum operation means 10 asks for the presumed superposition noise spectrum 5 using the noise superposition voice spectrum time series 201 which is the output of the spectrum operation means 3, and outputs it to the noise spectrum removal means 4. Drawing 25 is drawing showing the relation between the noise superposition sound signal 200 inputted from the input edge 1, and an analysis frame. Drawing 26 is drawing showing the voice section T2, T four and the noise section T1, and T3.
[0005] Next, actuation is explained. The voice model 204 for collating is created using the voice data which is beforehand uttered in a quiet environment and does not have noise superposition, and is stored in the reference model memory 8. In the spectrum operation means 3, a power spectrum is calculated for every analysis frame F1 and F2, ..., Fi, ..., FI, and the noise superposition sound signal 200 inputted from the input edge 1 is outputted as noise superposition voice spectrum time series 201. With the noise spectrum removal means 4, it considers that the presumed superposition noise spectrum 5 is overlapped as a power spectrum of a noise to each noise superposition voice spectrum of this noise superposition voice spectrum time series 201, and noise rejection processing shown in a formula (1) is performed.
[0006]
[Equation 1]

$$S(\omega) = \max\{X(\omega) - \alpha \cdot N(\omega), 0\} \qquad (1)$$

[0007] Here, power [in / in S (omega) / the frequency omega of a noise rejection voice spectrum], power [in / in X (omega) / the frequency omega of a noise superposition voice spectrum], and N (omega) are the power in the frequency omega of a presumed superposition noise spectrum. alpha is a parameter showing extent which takes a forward value and removes a noise component, and it is adjusted so that recognition precision may be made into max. In addition, max {...} is a function which returns the largest value in the element in the parenthesis divided with the comma.
[0008] To each noise rejection voice spectrum of the noise rejection voice spectrum time series 202 which the noise spectrum removal means 4 outputted, the feature-vector operation means 6 performs feature-vector data processing, and changes it into the vector which shows the acoustical description used in speech recognition, such as an auto correlation coefficient and an LPC (Linear PredictiveCoding) cepstrum. This is given to all the noise rejection voice spectrum time series 202, and it outputs as noise rejection voice feature-vector time series 203.
[0009] The collating means 7 performs collating with the voice model 204 for collating stored in the reference model memory 8 to the noise rejection voice feature-vector time series 203 which the feature-vector operation means 6 outputted, and outputs the category of the voice model 204 for collating which gives maximum likelihood as a recognition result 9.

[0010] Moreover, as the average spectrum operation means 10 is shown in drawing 26 by considering as an input the noise superposition voice spectrum time series 201 which is the output of the spectrum operation means 3 The section which is not voice in the noise superposition voice spectrum time series 201 For example, one or more noise superposition voice spectrums which hit at the noise section T1 in front of the voice section, the noise section T2 produced at the pause section under voice utterance are averaged for every frequency, and it outputs as a presumed superposition noise spectrum 5.

[0011] After considering that the noise sections T1 in front of voice utterance etc. or the average power spectrum of T3 was overlapped on the voice section T2 of the noise superposition voice spectrum time series 201 which it is as a result of the analysis of a spectrum of the noise superposition sound signal 200 inputted, or each noise superposition voice spectrum of T four and removing a noise component on a power spectrum by the above actuation, collating processing with a noise-less collating model is performed, and a recognition result is obtained.

[0012]

[Problem(s) to be Solved by the Invention] Since conventional equipment is constituted as mentioned above, when the difference of the noise sections T1 in front of voice utterance etc., the average power spectrum of T3, and the power spectrum of the noise superimposed on the voice of the actual voice section T2 and T four is small (i.e., when fluctuation of an ambient noise is small), it operates comparatively good. however, wearing of the voice-input microphone whose voice recognition unit user is an audio input edge -- it is working or a noise source is a migration object -- etc. -- the case where the distance of the input edge of a sound signal and a noise source carries out aging, and the ambient noise are unsteady, when fluctuation is large, the error of the average power spectrum obtained from just before voice utterance etc. and the power spectrum of the noise actually superimposed on voice became large, and there was a problem that where of the recognition engine performance deteriorates.

[0013] This invention was made in order to solve the above-mentioned problem, for every noise superposition voice spectrum of the noise superposition voice spectrum time series 201, using the noise-less voice model showing the noise model showing a noise signal, and noise-less voice, is asking for a presumed superposition noise spectrum, and aims at coping with the recognition performance degradation by fluctuation of an ambient noise, or distance fluctuation with a noise source and a voice input microphone.

[0014]

[Means for Solving the Problem] The noise model memory the bottom voice recognition unit of the noise concerning this invention remembers a noise model to be, The voice model memory which memorizes a noise-less voice model, and the reference model memory which memorizes the voice model for collating, A sonagraphy means to input a noise superposition sound signal, to perform sonagraphy and to output noise superposition voice feature-vector time series, The noise model memorized by noise model memory to the noise superposition voice feature-vector time series which is the output of said sonagraphy means, A superposition noise presumption means to perform superposition noise presumption processing and to output a presumed superposition noise spectrum using the noise-less voice model memorized by voice model memory, A spectrum operation means to input a noise superposition sound signal, to perform a analysis of a spectrum, and to output noise superposition voice spectrum time series, A noise spectrum removal means to remove the spectrum component of a noise using the presumed superposition noise spectrum which is the output of said superposition noise presumption means to the noise superposition voice spectrum time series which is the output of said spectrum operation means, A feature-vector operation means to calculate a feature vector and to output noise rejection voice feature-vector time series from the noise rejection voice spectrum time series which is the output of said noise spectrum removal means, Collating with the voice model for collating memorized by reference model memory is performed to the noise rejection voice feature-vector time series which is the output of said feature-vector operation means, and it is characterized by consisting of collating means to output the voice model for collating with which likelihood becomes high most as a recognition result.

[0015] Said superposition noise presumption means considers as an input the noise superposition voice feature-vector time series which is the output of a sonagraphy means. A presumed SN ratio operation means to calculate the presumed SN ratio of each noise superposition voice feature vector using the noise model memorized by noise model memory and the noise-less voice model memorized by voice model memory, According to the presumed SN ratio which is the output of said presumed SN ratio operation means, composition with the noise-less voice model memorized by the noise model memorized by noise model memory and voice model memory is performed. A noise superposition voice model generation means to generate a noise superposition voice model, and the noise superposition voice model which is the output of said noise superposition voice model generation means, A likelihood operation means to perform collating with the noise superposition voice feature vector set as the object of a presumed SN ratio operation in said presumed SN ratio operation means, and to output as collating data in quest of likelihood, A weight data operation means to calculate the weight data to the combination of three persons of each noise superposition voice feature vector, a noise model, and a noise-less voice model using the collating data outputted from said likelihood operation means, It is characterized by consisting of noise spectrum generation means to generate a presumed superposition noise spectrum, using the weight data outputted from said weight data operation means, the presumed SN ratio which is the output of said presumed SN ratio operation means, noise superposition voice feature-vector time series, and a noise model.

[0016] It is characterized by using the feature vector corresponding to one or more a noise spectrum and each noise spectrum as a noise model memorized by said noise model memory.

[0017] As a noise-less voice model memorized by said voice model memory, it is characterized by using one or more noise-less voice feature vectors.

[0018] It is characterized by using the model which connected the model of syllable mutually as a noise-less voice model memorized by said voice model memory.

[0019] While said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation means is characterized by choosing what has the highest likelihood and calculating weight data out of the collating data which said likelihood operation means outputted.

[0020] While said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation means is characterized by choosing in order two or more things which have high likelihood from the collating data which said likelihood operation means outputted, and calculating weight data.

[0021] While said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation means Two or more things which have high likelihood are chosen in order from the collating data which said likelihood operation means outputted, weighting addition of the likelihood obtained with the same noise model is carried out, and it is characterized by calculating weight data in quest of a noise model with the largest value as a result of this weighting addition.
[0022] The noise model memory the bottom speech recognition approach of the noise concerning this invention remembers a noise model to be, The sonagraphy process which is equipped with the voice model memory which memorizes a noise-less voice model, and the reference model memory which memorizes the voice model for collating, inputs a noise superposition sound signal, performs sonagraphy, and outputs noise superposition voice feature-vector time series, The noise model memorized by noise model memory to the noise superposition voice feature-vector time series which is the output of said sonagraphy process, The superposition noise presumption process which performs superposition noise presumption processing and outputs a presumed superposition noise spectrum using the noise-less voice model memorized by voice model memory, The spectrum operation process which inputs a noise superposition sound signal, performs a analysis of a spectrum, and outputs noise superposition voice spectrum time series, The noise spectrum removal process of removing the spectrum component of a noise using the presumed superposition noise spectrum which is the output of said superposition noise presumption process to the noise superposition voice spectrum time series which is the output of said spectrum operation process, The feature-vector operation process which calculates a feature vector from the noise rejection voice spectrum time series which is the output of said noise spectrum removal process, and outputs noise rejection voice feature-vector time series, Collating with the voice model for collating memorized by reference model memory is performed to the noise rejection voice feature-vector time series which is the output of said feature-vector operation process, and it is characterized by consisting of collating processes which output the voice model for collating with which likelihood becomes high most as a recognition result.
[0023] Said superposition noise presumption process considers as an input the noise superposition voice feature-vector time series which is the output of a sonagraphy process. The presumed SN ratio operation process of calculating the presumed SN ratio of each noise superposition voice feature vector using the noise model memorized by noise model memory and the noise-less voice model memorized by voice model memory, According to the presumed SN ratio which is the output of said presumed SN ratio operation process, composition with the noise-less voice model memorized by the noise model memorized by noise model memory and voice model memory is performed. The noise superposition voice model generation process which generates a noise superposition voice model, and the noise superposition voice model which is the output of said noise superposition voice model generation process, The likelihood operation process which performs collating with the noise superposition voice feature vector set as the object of a presumed SN ratio operation in said presumed SN ratio operation process, and is outputted as collating data in quest of likelihood, The weight data operation process of calculating the weight data to the combination of three persons of each noise superposition voice feature vector, a noise model, and a noise-less voice model using the collating data outputted from said likelihood operation process, It is characterized by consisting of noise spectrum generation processes which generate a presumed superposition noise spectrum using the weight data outputted from said weight data operation process, the presumed SN ratio which is the output of said presumed SN ratio operation process, noise superposition voice feature-vector time series, and a noise model.
[0024] While said likelihood operation process calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation process is characterized by choosing what has the highest likelihood and calculating weight data out of collating data.
[0025] Said likelihood operation process is characterized by for said weight data operation process choosing in order two or more things which have high likelihood from collating data, and calculating weight data while it calculates and outputs collating data for every combination of a noise model and a voice model.
[0026] Said weight data operation process chooses in order two or more things which have high likelihood from collating data, and said likelihood operation process carries out weighting addition of the likelihood obtained with the same noise model, and is characterized by to calculate weight data in quest of a noise model with the largest value as a result of this weighting addition while it calculates and outputs collating data for every combination of a noise model and a voice model.
[0027]
[Embodiment of the Invention]
Gestalt 1. drawing 1 of operation is a speech recognition structure-of-a-system Fig. concerning this invention. The voice recognition system 60 is equipped with the display unit 61, a keyboard 62, a mouse 63, the mouse putt 64, the system unit 65, the voice input microphone 66, and the voice recognition unit 100. As shown in drawing 1 , the voice recognition system of this invention recognizes the voice inputted from the voice input microphone 66 with a voice recognition unit 100, transmits the recognized voice to a system unit 65, and displays it on the display unit 61 as an alphabetic character. However, as long as the voice recognition system concerning this invention is a system by which the voice recognition unit 100 which it does not need to be used together with a personal computer or a workstation in this way, and is described below is used, it may be the thing of what kind of format. For example, a tape recorder may be used as an input device and you may make it input the voice data from a network instead of the voice input microphone 66. Moreover, the data to input may be analog data and may be digital data. Moreover, although a voice recognition unit 100 may exist with the independent case, it may be the case where it may be dedicated to the interior of a system unit 65, and exists as a part of system board of other measurement machines or a calculating machine. Moreover, you may make it make data retrieval, processing, and measurement perform not only based on when displaying a recognition result as an alphabetic character, but a recognition result.
[0028] Drawing 2 is the block diagram showing the configuration of the gestalt of 1 operation of the voice recognition unit 100 under the noise concerning this invention. In drawing, 11 performs sonagraphy to the noise superposition sound signal 200 inputted from the input edge 1. A sonagraphy means to output the noise superposition voice feature-vector time series 120, the voice model memory which has memorized the noise-less voice model with which 13 expresses noise-less voice, The noise model 121 memorized by noise model memory to the noise model memory which has memorized the noise model with which 14 expresses a noise, and the noise superposition voice feature-vector time series 120 to which 15 was outputted from the sonagraphy means 11, It is a superposition noise presumption means to perform superposition noise presumption and to output the presumed superposition noise spectrum 5 using the noise-less voice model 122 memorized by voice model memory. Other components are the same as the conventional example.

[0029] Drawing 3 is the block diagram showing the configuration of the gestalt of 1 operation of the superposition noise presumption means 15 in the voice recognition unit 100 under the noise concerning this invention. The noise-less voice model 122 memorized by the voice model memory 13 in drawing to each noise superposition voice feature vector of the noise superposition voice feature-vector time series 120 whose 16 is the output of the sonagraphy means 11, A presumed SN ratio operation means to calculate presumed SN ratio 130 using the noise model 121 memorized by the noise model memory 14, 17 follows the presumed SN ratio data which are the output of the presumed SN ratio operation means 16. A noise superposition voice model generation means to perform composition of the noise-less voice model 122 and the noise model 121, and to generate the noise superposition voice model 131, 18 performs collating with the noise superposition voice model 131 and the noise superposition voice feature-vector time series 120 which were generated. It asks for the matching data and likelihood of each noise superposition voice feature vector and the noise superposition voice model 131. A likelihood operation means to output as collating data 132, and 19 consider the collating data 132 which are the output of the likelihood operation means 18 as an input. A maximum likelihood selection mold weight data operation means to create the likelihood maximum criteria weight data 133, and 20 The noise superposition voice feature-vector time series 120, The presumed SN ratio data which are the output of a presumed SN ratio operation means, and the weight data 133 which are the output of the maximum likelihood selection mold weight data operation means 19, It is a noise spectrum generation means to generate the presumed superposition noise spectrum 5 using the noise model 121 memorized by the noise model memory 14.

[0030] Next, actuation is explained. First, an auto correlation coefficient and a normalization auto correlation coefficient are explained using drawing 4 . Drawing 4 is drawing showing the analysis frame F1. In the analysis frame F1, A samplings are performed and the value of the feature vector in each sampling is set to B (1), B (2), ..., B (a), ..., B (A). An auto correlation coefficient integrates value [of its own feature vector] B (a), and the value B of a feature vector (a+k) which separated only k from itself, and is called for by taking the total. That is, auto correlation coefficient z (k) is carried out like a formula (2), and is called for.

[0031]

[Equation 2]

$$z(k)=\sum_{a=1}^{A-k}B(a)\cdot B(a+k) \qquad (2)$$

[0032] Here, k expresses a dimension and k= 0 shows zero-order origin. That is, z (0) is value B(a)2 of a feature vector, as the auto correlation coefficient of zero-order origin is shown and a zero-order auto correlation coefficient is shown in a formula (3). It is totaled.

[0033]

[Equation 3]

$$z(0)=\sum_{a=1}^{A-0}B(a)\cdot B(a+0)=\sum_{a=1}^{A}B(a)^2 \qquad (3)$$

[0034] This zero-order auto correlation coefficient is a value which shows power. And as shown in a formula (4), it is the normalization auto correlation coefficient zn (k) which normalized auto correlation coefficient z (k) by power.

[0035]

[Equation 4]

$$zn(k)=\frac{z(k)}{z(0)} \qquad (4)$$

[0036] And what vectorized as an example the normalization auto correlation coefficient called for by doing in this way from zero-order to the 15th order is normalization autocorrelation charge numerical vector Z shown in a formula (5).

[0037]

[Equation 5]

$$Z=\{zn(0),zn(1),\cdots,zn(15)\} \qquad (5)$$

[0038] In addition, an auto correlation coefficient may carry out the inverse Fourier transform of the power spectrum in the analysis frame F1, and may ask for it. Although the case where it asked for a normalization autocorrelation charge numerical vector from a frame F1 was explained when shown in drawing 4 , it can ask for a normalization autocorrelation charge numerical vector similarly about noise-less voice and a noise.

[0039] Below, concrete actuation of the gestalt of this operation is explained. it is shown in drawing 5 -- as -- a power spectrum with noise-less beforehand typical voice, for example, a vowel, "**" -- "-- it is -- " -- "-- obtaining -- " -- "-- obtaining -- " -- "-- " -- it asks for two or more (M individual) power spectrums of various consonants, and the normalization autocorrelation charge numerical vector which corresponds, respectively is stored in the voice model memory 13 as a noise-less voice model 122. That is, one or more noise-less voice feature vectors are memorized as a noise-less voice model memorized by voice model memory. The m-th normalization autocorrelation charge numerical vector is set to Sm (1 <=m<=M). In noise-less voice, it is considered that the normalization autocorrelation charge numerical vector which hits each representation point is observed by same probability.

[0040] Moreover, as shown in drawing 6 , it asks for a typical power spectrum [two or more (N individual)] also about a noise, and stores in the noise model memory 14 by using the normalization autocorrelation charge numerical vector and normalization power spectrum corresponding to it as the noise model 121. That is, the description corresponding to one or more a noise spectrum and each spectrum is memorized as a noise model memorized by noise model memory. A normalization power spectrum can carry out the Fourier transform of the normalization auto correlation coefficient, and can ask for it. Therefore, when allowances of enough are in the processing time, it is not necessary to calculate a normalization power spectrum beforehand, and whenever there is need, the Fourier transform is performed from a normalization autocorrelation charge numerical vector, and you may make it ask for a normalization

power spectrum. Here, the n-th normalization autocorrelation charge numerical vector is set to Vn (1 <=n<=N), and a normalization power spectrum is set to Wn (1 <=n<=N). Moreover, in a noise, it is considered that the normalization autocorrelation charge numerical vector which hits each representation point is observed by same probability.

[0041] About the noise superposition sound signal 200 inputted from the input edge 1, to each analysis frames F1 and F2 set as the object of the analysis-of-a-spectrum processing in the spectrum operation means 3, ..., Fi, ..., FI, the sonagraphy means 11 performs sonagraphy and outputs it as noise superposition voice feature-vector time series 120. As shown in drawing 7 , the normalization autocorrelation charge numerical vector of the noise superposition voice feature vector of the i-th frame is set to Xi (1 <=i<=I and I are a frame number). The presumed SN ratio operation means 16 asks for presumed SN ratio Ri (m, n) (1 <=i<=I, 1 <=m<=M, 1 <=n<=N) like a formula (6) from the noise superposition voice feature-vector time series 120 using the noise-less voice model 122 memorized by the voice model memory 13 and the noise model 121 memorized by the noise model memory 14, as shown in drawing 8 .

[0042]

[Equation 6]

$$Ri(m,n)=\frac{\sum_{k=-K}^{K} Am(k)\cdot Vn(k)-\sum_{k=-K}^{k} Am(k)\cdot Xi(k)}{\sum_{k=-K}^{K} Am(k)\cdot Xi(k)-\sum_{k=-K}^{K} Am(k)\cdot Sm(k)} \quad (6)$$

[0043] It is a presumed SN ratio when, as for expressing with (k), m-th normalization autocorrelation charge numerical vector Sm of the noise-less voice model 122 and the n-th normalization autocorrelation charge numerical vector Vn of the noise model 121 are used for Ri (m, n) to the noise superposition voice feature vector of the i-th frame to the k-dimensional component of each vector here. Moreover Am is the maximum ** parameter called for from normalization autocorrelation charge numerical vector Sm of the noise-less voice model 122.

[0044] As shown in drawing 9 , the noise superposition voice model generation means 17 The noise-less voice model 122 memorized by the voice model memory 13 according to the presumed SN ratio data Ri (m, n) (1 <=i<=I, 1 <=m<=M, 1 <=n<=N) which are the output of the presumed SN ratio operation means 16, The noise model 121 memorized by the noise model memory 14 is used. The noise superposition voice model Ti (m, n) (1 <=i<=I, 1 <=m<=M, 1 <=n<=N) like a formula (7) It generates as the weighting sum according to the presumed SN ratio of the normalization autocorrelation charge numerical vector of the noise-less voice model 122, and the normalization autocorrelation charge numerical vector of the noise model 121.

[0045]

[Equation 7]

$$Ti(m,n)=\frac{Ri(m,n)}{1+Ri(m,n)}\cdot Sm+\frac{1}{1+Ri(m,n)}\cdot Vn \quad (7)$$

[0046] As shown in drawing 10 , from the noise superposition voice feature-vector time series 120, the likelihood operation means 18 performs collating with the noise superposition voice model 131 which is the output of the noise superposition voice model generation means 17, and asks for likelihood Li (m, n) (1 <=i<=I, 1 <=m<=M, 1 <=n<=N) like a formula (8).

[0047]

[Equation 8]

$$Li(m,n)=d\left(Xi,Ti(m,n)\right)^{-1} \quad (8)$$

[0048] Here, d (*, *) is the Euclidean distance when expressing a suitable distance defined between two normalization autocorrelation charge numerical vectors, for example, changing each normalization autocorrelation charge numerical vector into an LPC cepstrum vector by LPC analysis etc.

[0049] The maximum likelihood selection mold weight data operation means 19 asks for the sequence data (Pi, Qi) of the noise superposition voice model 131 made into likelihood max from the noise superposition voice feature-vector time series 120 first using the likelihood data which are the output of a likelihood operation means. That is, for example, it asks for sequence data which fill a formula (9) and a formula (10).

[0050]

[Equation 9]

$$\sum_{i=1}^{I} Li(Pi,Qi) > \sum_{i=1}^{I} Li(m,n) \quad m\neq Pi, n\neq Qi \quad (9)$$

[0051]

[Equation 10]

$$\prod_{i=1}^{I} Li(Pi,Qi) > \prod_{i=1}^{I} Li(m,n) \quad m\neq Pi,\ n\neq Qi \quad (10)$$

[0052] How to ask for the above-mentioned sequence data is explained using drawing 11 . Drawing 11 shows how to ask for sequence data (Pi, Qi) in case the likelihood Li (m, n) shown in drawing 10 consists of M piece xN individual =2x2 and the 1st and 2nd analysis frame (i= 1, i= 2) exists (I= 2). When it is asked for likelihood to show drawing 11 , a formula (5) calculates total of all the combination of the likelihood obtained from the 1st and 2nd analysis frame, and selects that from which the total serves as max. When shown in

drawing 11 , 5+4=9 become max. Therefore, the case of likelihood L1(1 1)+L2(2 2) =9 serves as max. Therefore, it asks for = (P1, Q1) (1 1) and = (P2, Q2) (2 2) as sequence data.

[0053] Next, according to this sequence data, it asks for the weight data Gi (m, n) like a formula (11).

[0054]

[Equation 11]

$$Gi(m,n)=\begin{cases}1, & m=Pi, n=Qi\\ 0, & m\neq Pi, n\neq Qi\end{cases} \qquad (11)$$

[0055] Drawing 12 is drawing showing the weight data called for from the example shown in drawing 11 . Only the weight shown by sequence data (1 1) is set to 1 to the 1st analysis frame (i= 1), and other values are set to 0. Moreover, the weight data corresponding to sequence data (2 2) also in the case of the 2nd analysis frame (i= 2) are set to 1, and other weight data are set to 0.

[0056] The noise spectrum generation means 20 generates the presumed superposition noise spectrum 5 using the power spectrum of the noise model 121 remembered to be the weight data which are the output of the maximum likelihood selection mold weight data operation means 19, the presumed SN ratio data which the presumed SN ratio operation means 16 calculated, and the noise superposition voice feature-vector time series 120 by the noise model memory 14. first, about the combination of i whose weight data Gi (m, n) are not 0 (that is, it is 1), and m and n The presumed SN ratio data Ri (m, n) and the noise power data Ui in the noise superposition voice feature vector of the i-th frame of the noise superposition voice feature-vector time series 120, It asks for the superposition noise power spectrum Zi (m, n) by the formula (12) using the normalization power spectrum Wn memorized by the noise model memory 14. Thereby, the facies of a power spectrum of Zi (m, n) are the same as that of the normalization power spectrum Wn of the noise model 121, and it serves as a power spectrum with the noise power in the i-th noise superposition voice feature vector according to the presumed SN ratio data Ri (m, n).

[0057]

[Equation 12]

$$Zi(m,n)=\frac{Ui}{1+Ri(m,n)}\cdot Wn \qquad (12)$$

[0058] When shown in drawing 12 , the combination of i whose weight data are 1, and m and n is the case of i= 1, m= 1, and n= 1, and two cases of i= 2, m= 2, and n= 2. Therefore, a superposition noise power spectrum is calculated about two cases, Z1 (1 1) and Z2 (2 2). Since weight data are 0 in the case of others, count of a superposition noise power spectrum is not performed.

[0059] Next, the weighting average of the superposition noise power spectrum Zi (m, n) by the weight data Gi (m, n) is performed like a formula (13) about all the combination of m and n to every i, and obtained Yi is outputted as a presumed superposition noise spectrum 5.

[0060]

[Equation 13]

$$Yi=\frac{\sum_{1\leq m\leq M,1\leq n\leq N} Gi(m,n)\cdot Zi(m,n)}{\sum_{1\leq m\leq M,1\leq n\leq N} Gi(m,n)} \qquad (13)$$

[0061] Count of the above-mentioned weighting average has semantics especially in the gestalt 2 of operation mentioned later, and the gestalt 3 of operation, and in the gestalt 1 of this operation, since weight data are 1, the superposition noise power spectrum Zi (m, n) is outputted as it is. In the case of the example shown in drawing 12 , it is Y1=Z1 (1 1).

Y2=Z2(2,2)

It is outputted by carrying out. After the presumed superposition noise spectrum 5 is obtained, the same recognition collating processing as the conventional example is performed, and a recognition result is obtained.

[0062] By the above processings, the noise superposition voice feature-vector time series 120 is received. Since a likelihood operation with the noise-less voice model 122 in consideration of the noise which the noise model 121 expresses being overlapped is performed and power spectrum presumption of the superposition noise by the likelihood maximum criteria is made for every analysis frame of noise superposition input voice, The presumed superposition noise spectrum which cannot be easily influenced by the superposition noise by distance fluctuation with an unsteady noise environment, and a voice input microphone and a noise source of fluctuation will be obtained, and degradation of the recognition engine performance can be suppressed.

[0063] Moreover, the normalization autocorrelation charge numerical vector corresponding to the typical power spectrum of noise-less voice considers that it is observed by same probability in noise-less voice. Having constituted the noise-less voice model 122 and a noise are similarly regarded as the normalization autocorrelation charge numerical vector corresponding to the typical power spectrum of a noise being observed by same probability in a noise. By having constituted the noise model 121, the decision of the sequence data which fill a formula (9) and a formula (10) can substitute finding each about i the combination of m and n with which a formula (14) is filled.

[0064]

[Equation 14]

$$Li(Pi,Qi) > Li(m,n) \quad m\neq Pi, n\neq Qi \qquad (14)$$

[0065] The decision approach of the above-mentioned sequence data is concretely explained using drawing 11 . The above-mentioned formula (14) should just find the maximum likelihood on each analysis frame i of every. Therefore, in the case of i= 1 shown in drawing 11 , the maximum likelihood 5 should just be chosen out of four likelihood 5, 1, 4, and 2. Moreover, in the case of i= 2, the maximum

likelihood 4 should just be chosen out of four likelihood 3, 1, 2, and 4. In this example, although the same likelihood as drawing 11 mentioned above is chosen and (2, 2) are determined as the sequence data same as a result as the case of drawing 11 (1 1), different sequence data depending on the case are determined. Thus, if a formula (14) is followed, since maximum likelihood can be determined for every analysis frame, it is not necessary to wait to calculate the likelihood of all analysis frames like [at the time of following a formula (9) and a formula (10)]. Thereby, for every analysis frame of noise superposition input voice, a presumed superposition noise spectrum can be presumed and improvement in the speed according [processing to the feature-vector operation means 6] to pipeline processing etc. is attained.

[0066] Below, the procedure of the gestalt of this operation is indicated collectively.

(1) Ask for a presumed SN ratio about all the combination of a noise model and a noise-less voice model from the analysis frame of the arbitration of input voice.

(2) Make a noise model superimpose on a noise-less voice model about each combination according to the obtained presumed SN ratio.

(3) Find the distance of the analysis frame of input voice, and the noise superposition voice called for by (2), and determine the combination of the noise model which gives the minimum distance, and a noise-less voice model.

Noise power is calculated from the power of the presumed SN ratio by the combination determined by (4) and (3), and the applicable analysis frame of input voice.

(5) The power spectrum corresponding to the noise model of the combination determined as the analysis frame of input voice by (3) considers that it superimposes by the power calculated by (4), and carries out spectrum subtraction processing.

(6) (1) - (5) is performed about all the analysis frames of input voice.

By the above processing, the noise model selection of a superposition noise and power count by the likelihood maximum criteria with the noise-less voice model which took superposition of a noise into consideration for every analysis frame of input voice are made, and noise rejection which cannot be easily influenced by the unsteady noise or the presumed SN ratio of fluctuation is performed.

[0067] The noise-less voice model 122 in this invention is not limited to the normalization autocorrelation charge numerical vector corresponding to two or more above-mentioned typical power spectrums, and the model which connected mutually the model of syllable which appears in Japanese altogether may be used for it. the normalization autocorrelation charge numerical vector corresponding to the mean vector [in / when HMM (Hiden Marcov Model) of a continuous-distribution mold is used as a model of syllable / HMM] -- every -- it asks for every mean vector of HMM, and stores in voice model memory as a noise-less voice model 122. The presumed SN ratio operation in a formula (6) is performed using this normalization autocorrelation charge numerical vector. Generation of a noise superposition voice model carries out sonagraphy of the autocorrelation charge numerical vector obtained by the weighting sum which follows the presumed SN ratio of the normalization autocorrelation charge numerical vector of the noise-less voice model 122, and the normalization autocorrelation charge numerical vector of the noise model 121 by the formula (7), and is performed by replacing the analysis result with the mean vector of HMM. The likelihood of a formula (8) asks for the sequence data (Pi, Qi) in the maximum likelihood selection mold weight data operation means 19 instead of a formula (9) and a formula (10) by the Viterbi operation in HMM using the output probability in HMM. The precision as a noise-less voice model 122 showing noise-less voice improves by this, the presumed precision of a superposition noise is improved, and bottom speech recognition of the noise with a more high precision is realized.

[0068] Gestalt 2. drawing 13 of operation is the block diagram showing the configuration of the gestalt of 1 operation of the superposition noise presumption means 15 in this invention. The difference with the gestalt 1 of operation is the point of using two or more candidate concomitant use mold weight data operation means 21, instead of a maximum likelihood selection mold weight data operation means.

[0069] Next, actuation is explained. The likelihood data which are the output of the likelihood operation means 18 are first used for two or more candidate concomitant use mold weight data operation means 21. For every noise superposition voice feature vector of the noise superposition voice feature-vector time series 120 Sorting is performed for likelihood Li (m, n) about 1 <=m<=M and 1 <=n<=N, and it asks for the sequence data (Pi (j), Qi (j)) of the combination (m, n) of m and n which give the likelihood to descending of likelihood. j -- ranking -- expressing -- Pi -- (j {1, 2, ..., M}) and Qi -- it is (j {1, 2, ..., N}).

[0070] Next, according to this sequence data, it asks for the weight data Gi (m, n) like a formula (15). H is the number of the likelihood high order candidates who use for generation of a presumed superposition noise spectrum among a formula.

[0071]

[Equation 15]

$$Gi(m,n)=\begin{cases} 1/H, & (m,n)\in\{(Pi(j),Qi(j))\mid 1\le j\le H\} \\ 0, & (m,n)\notin\{(Pi(j),Qi(j))\mid 1\le j\le H\} \end{cases} \quad (15)$$

[0072] Drawing 14 is drawing showing the example of the gestalt of this operation. Drawing 14 shows the weight data in the case of the H= 2 number of likelihood high order candidates, when likelihood as shown in drawing 11 is calculated. In the case of the 1st analysis frame (i= 1) shown in drawing 11 , it will be set to 5, 4, 2, and 1 if it arranges to descending of likelihood. Weight data are assigned [as opposed to / moreover / two likelihood 5 and 4] every [2 / 1/] equally. Moreover, if the likelihood of the 2nd analysis frame (i= 2) is arranged in descending, it will be set to 4, 3, 2, and 1, and weight data will be given for likelihood every [2 / 1/] to 4 and 3. Thus, when weight is assigned every [2 / 1/], the average of two superposition noise power spectrums is calculated by the formula (13) mentioned above, and it is outputted as a presumed superposition noise spectrum Yi. The presumed superposition noise spectrum which used together the combination of two or more m and n with large likelihood by this will be generated, degradation of the precision of the presumed superposition noise spectrum by the error of the noise model 121 and an actual noise and the error of the noise-less voice model 122 and that of actual voice can be suppressed, and recognition precision also improves.

[0073] The operation expression of weight data is not limited to a formula (15), and seems moreover, to enlarge weight in proportion to likelihood, as shown in drawing 15 .

[0074] Gestalt 3. drawing 16 of operation is the block diagram showing the configuration of the gestalt of 1 operation of the

superposition noise presumption means 15 in this invention. The difference with the gestalt 1 of operation is the point of using the average likelihood operation mold weight data operation means 22, instead of a maximum likelihood selection mold weight data operation means.

[0075] Next, actuation is explained. the sequence Pin of m which the average likelihood operation mold weight data operation means 22 carries out sorting of the likelihood Li (m, n) about 1 <=m<=M to n of arbitration first using the likelihood data which are the output of the likelihood operation means 18 for every noise superposition voice feature vector of the noise superposition voice feature-vector time series 120, and gives the likelihood to descending of likelihood -- it asks for (j {1, 2, ..., M}). j expresses ranking. Subsequently, it asks by averaging Li (Pin (j), n) which fills j<=C for the average likelihood Ein in every n. Here, C is the number of the likelihood high order candidates who use for count of average likelihood. In 1 <=n<=N, if n which gives the maximum of Ein is set to nmax, it will ask for the weight data Gi (m, n) like a formula (16).

[0076]

[Equation 16]

$$Gi(m,n)=\begin{cases} 1/C, & m\in\{Pin(j)\mid 1\le j\le C\}, n=n\max \\ 0, & others \end{cases} \quad (16)$$

[0077] The concrete calculation approach of weight data in the gestalt 3 of this operation is explained using drawing 17 . Drawing 17 shows the computed likelihood. Here, several C of a likelihood high order candidate is set to C= 2. In the case of n= 1, two high order candidates are 5 and 3, and the average is 4. In the case of n= 2, high order candidates are 4 and 2 and the average is 3. Similarly, in the case of n= 3, in the case of 3.5 and n= 4, the average is set to 3 by the average. Therefore, as for an average value, it is determined that the case of n= 1 serves as max and uses the noise model of n= 1. And as shown in drawing 18 , the weight data which are a likelihood high order candidate's several C inverse number are calculated. It seems that the operation expression of weight data is not limited to a formula (16) although the case where weight is only equally set to one half is shown, and weight may be here enlarged in proportion to likelihood although not illustrated.

[0078] Thereby, in combination with the noise-less voice model 122, it asks for average likelihood every noise model 121, using two or more what has high likelihood, and in order to choose the noise model 121 which gives the maximum average likelihood, degradation of the precision of the presumed superposition noise spectrum by the error of the noise model 121 and an actual noise and the error of the noise-less voice model 122 and that of actual voice can be suppressed, and recognition precision also improves.

[0079] The example of an experiment

(1) The specified speaker dispersion word recognition experiment based on the piece HMM of a phoneme of a continuous-distribution mold estimated the noise rejection technique stated with the gestalt 1 of experiment condition operation. Experiment conditions are shown in drawing 19 . The discrete word recognition experiment by the piece HMM of a phoneme is conducted to 100 name of a place voice which made the unsteady noise in works superimpose on a PINK noise list. Noise superposition voice was obtained by adding noise data to noise-less voice data on a calculating machine. The presumed SN ratio is defined as the average power of the voice data except the noise section by the average power ratio of noise data. It asked for the voice model by performing clustering by the LBG (Linde, Buzo, Gray) algorithm, and averaging the normalization auto correlation coefficient of each cluster about the voice section of the voice for study. From each noise data of all, the noise model of each noise performed clustering by the LBG algorithm, and asked for the average of the normalization auto correlation coefficient of each category, and the average of a power spectrum. Both the interval scales in clustering are the Euclidean distances of an LPC cepstrum. the spectrum subtraction (SS) which makes the average of the power spectrum of ten frames in front of voice a presumed superposition noise as the conventional technique at a recognition experiment using 100 gazetteers -- it compared with law.

(2) evaluation with PINK noise superposition voice -- recognition evaluation by the voice data on which the steady PINK noise was made to superimpose was performed. A presumed SN ratio is 10dB. As for one cluster and a voice model, the noise model used the centroid at the time of 64 clusters. The result of having searched for the incorrect recognition rate when changing the noise rejection multiplier in spectrum subtraction (SS) about each speaker (TH, TS) is drawing 20 . SSs are a conventional method and the noise rejection technique which EST-SS stated with the gestalt 1 of operation among drawing. In the noise rejection technique stated with the gestalt 1 of operation, it is based neither on a speaker nor a removal multiplier, but it is almost the same as the conventional method, or it turns out that the removal multiplier from which the recognition engine performance's good a little being obtained and incorrect recognition serve as min is 2.0. It was checked that the noise rejection technique stated from this with the gestalt 1 of operation to the voice data which the steady PINK noise superimposed operates good.

(3) The recognition experiment was conducted using the voice data on which the unsteady noise recorded in the evaluation with unsteady noise superposition voice, next inspection Rhine of an auto factory was made to superimpose. A noise model is a centroid with four clusters, and the voice model is as common as a previous experiment. Drawing 21 is as a result of the recognition experiment to the unsteady noise superposition voice data whose presumed SN ratio is 10dB. The shaft is the same as drawing 20 in the notation list in drawing. Though it is the same presumed SN ratio as compared with the experimental result of the point using a PINK noise, it is not based on a method, a speaker, and a removal multiplier, but incorrect recognition is increasing, and the degradation by the unsteady noise is remarkable. The noise rejection technique stated with the gestalt 1 of operation is based neither on a speaker nor a removal multiplier, but the recognition precision exceeding a conventional method type is acquired, and the effectiveness of the noise rejection technique stated with the gestalt 1 of operation is clear also in an unsteady noise. Next, the recognition experiment when changing the presumed SN ratio of voice data was conducted. A result is shown in drawing 22 . An axis of abscissa is a presumed SN ratio. The removal multiplier was set to 2.0 from the previous experimental result. Here, the recognition engine performance when not performing noise rejection processing at all is also shown as comparison data. It turns out that it is not based on a presumed SN ratio, but the noise rejection technique stated with the gestalt 1 of operation is operating effectively.

(4) It connected, and as mentioned above, superposition noise presumption was carried out using a noise model and a noise-less voice model as the recognition technique of noise superposition voice, and it experimented in the technique removed on a power spectrum.

The effectiveness was checked as a result of the recognition experiment by the 100 name-of-a-place voice data which made unsteady factory noise superimpose on a PINK noise list.

[0080]

[Effect of the Invention] Since this invention is constituted as explained above, it has the effectiveness which is indicated below.

[0081] As mentioned above, according to this invention, an error with the noise spectrum actually superimposed on the presumed superposition noise spectrum by fluctuation of an ambient noise etc. is small, and degradation of recognition precision is suppressed.

[0082] Moreover, according to this invention, the noise superposition voice feature-vector time series 120 is received. Since a likelihood operation with the noise-less voice model 122 in consideration of the noise which the noise model 121 expresses being overlapped is performed and power spectrum presumption of the superposition noise by the likelihood maximum criteria is made for every analysis frame of noise superposition input voice, The presumed superposition noise spectrum which cannot be easily influenced by the superposition noise by distance fluctuation with an unsteady noise environment, and a voice input microphone and a noise source of fluctuation will be obtained, and degradation of the recognition engine performance can be suppressed.

[0083] Moreover, according to this invention, in order to memorize the feature vector corresponding to one or more a noise spectrum and each spectrum in noise model memory as a noise model 121, the superposition noise by the unsteady ambient noise can be modeled with a sufficient precision, the precision of a presumed superposition noise spectrum improves, and degradation of recognition precision is suppressed.

[0084] Moreover, in order to memorize in voice model memory by using one or more noise-less voice feature vectors as the noise-less voice model 122 according to this invention, noise-less voice can be modeled with a sufficient precision, and the precision of a presumed superposition noise spectrum improves. Moreover, a presumed superposition noise spectrum can be presumed on the analysis frame of noise superposition input voice at **, and improvement in the speed according [processing to a feature-vector operation means] to pipeline processing etc. is attained.

[0085] Moreover, in order to memorize the model which connected the model of syllable mutually altogether in voice model memory as a noise-less voice model 122 according to this invention, the precision as a noise-less voice model 122 showing noise-less voice improves, the presumed precision of a superposition noise is improved, and bottom speech recognition of the noise with a more high precision is realized.

[0086] Moreover, according to this invention, a likelihood operation means calculates for every combination of the noise model 121 and a voice model, and since what has the highest likelihood is chosen from the collating data 132 to output and weight data are calculated, it can ask for a presumed superposition noise spectrum in the small amount of operations.

[0087] Moreover, according to this invention, a likelihood operation means calculates for every combination of the noise model 121 and a voice model, and in order to choose in order two or more things which have high likelihood from the collating data 132 to output and to calculate weight data, the precision fall of the presumed superposition noise spectrum by the error with the noise actually superimposed on the noise model 121 and the error of the noise-less voice model 122 and actual voice is controlled.

[0088] Moreover, according to this invention, a likelihood operation means calculates for every combination of the noise model 121 and a voice model. Two or more things which have high likelihood are chosen in order from the collating data 132 to output. In order to carry out weighting addition of the likelihood obtained with the same noise model 121 and to calculate weight data in quest of the noise model 121 with the largest value as a result of this weighting addition, The precision fall of the presumed superposition noise spectrum by the error with the noise actually superimposed on the noise model 121 and the error of the noise-less voice model 122 and actual voice is controlled.

[Translation done.]

CLAIMS

[Claim(s)]

[Claim 1] The noise model memory which memorizes a noise model, and the voice model memory which memorizes a noise-less voice model, The reference model memory which memorizes the voice model for collating, and a sonagraphy means to input a noise superposition sound signal, to perform sonagraphy, and to output noise superposition voice feature-vector time series, The noise model memorized by noise model memory to the noise superposition voice feature-vector time series which is the output of said sonagraphy means, A superposition noise presumption means to perform superposition noise presumption processing and to output a presumed superposition noise spectrum using the noise-less voice model memorized by voice model memory, A spectrum operation means to inpu a noise superposition sound signal, to perform a analysis of a spectrum, and to output noise superposition voice spectrum time series, A noise spectrum removal means to remove the spectrum component of a noise using the presumed superposition noise spectrum which is the output of said superposition noise presumption means to the noise superposition voice spectrum time series which is the output of said spectrum operation means, A feature-vector operation means to calculate a feature vector and to output noise rejection voice feature-vector time series from the noise rejection voice spectrum time series which is the output of said noise spectrum removal means, As opposed to the noise rejection voice feature-vector time series which is the output of said feature-vector operation means The bottom voice recognition unit of the noise characterized by consisting of collating means to output the voice model for collating with which collating with the voice model for collating memorized by reference model memory is performed, and likelihood becomes high most as a recognition result.

[Claim 2] Said superposition noise presumption means considers as an input the noise superposition voice feature-vector time series which is the output of a sonagraphy means. A presumed SN ratio operation means to calculate the presumed SN (single noise) ratio of each noise superposition voice feature vector using the noise model memorized by noise model memory and the noise-less voice model memorized by voice model memory, According to the presumed SN ratio which is the output of said presumed SN ratio operation means, composition with the noise-less voice model memorized by the noise model memorized by noise model memory and voice model memory is performed. A noise superposition voice model generation means to generate a noise superposition voice model, and the noise superposition voice model which is the output of said noise superposition voice model generation means, A likelihood operation means to perform collating with the noise superposition voice feature vector set as the object of a presumed SN ratio operation in said presumed SN ratio operation means, and to output as collating data in quest of likelihood, A weight data operation means to calculate the weight data to the combination of three persons of each noise superposition voice feature vector, a noise model, and a noise-less voice model using the collating data outputted from said likelihood operation means, The weight data outputted from said weight data operation means, and the presumed SN ratio which is the output of said presumed SN ratio operation means, The bottom voice recognition unit of the noise according to claim 1 characterized by consisting of noise spectrum generation means to generate a presumed superposition noise spectrum, using noise superposition voice feature-vector time series and a noise model.

[Claim 3] The bottom voice recognition unit of the noise according to claim 2 characterized by using the feature vector corresponding to one or more a noise spectrum and each noise spectrum as a noise model memorized by said noise model memory.

[Claim 4] The bottom voice recognition unit of the noise according to claim 2 characterized by using one or more noise-less voice feature vectors as a noise-less voice model memorized by said voice model memory.

[Claim 5] The bottom voice recognition unit of the noise according to claim 2 characterized by using the model which connected the model of syllable mutually as a noise-less voice model memorized by said voice model memory.

[Claim 6] It is the bottom voice recognition unit of the noise according to claim 2 characterized by for said weight data operation means choosing what has the highest likelihood from the collating data which said likelihood operation means outputted while said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, and calculating weight data.

[Claim 7] It is the bottom voice recognition unit of the noise according to claim 2 characterized by for said weight data operation means choosing in order two or more things which have the high likelihood out of the collating data which said likelihood operation means outputted while said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, and calculating weight data.

[Claim 8] While said likelihood operation means calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation means Two or more things which have high likelihood are chosen in order from the collating data which said likelihood operation means outputted. The bottom voice recognition unit of the noise according to claim 7 which carries out weighting addition of the likelihood obtained with the same noise model, and is characterized by calculating weight data in quest of a noise model with the largest value as a result of this weighting addition.

[Claim 9] The noise model memory which memorizes a noise model, and the voice model memory which memorizes a noise-less voice model, The sonagraphy process which is equipped with the reference model memory which memorizes the voice model for collating, inputs a noise superposition sound signal, performs sonagraphy, and outputs noise superposition voice feature-vector time series, The noise model memorized by noise model memory to the noise superposition voice feature-vector time series which is the output of said

sonagraphy process, The superposition noise presumption process which performs superposition noise presumption processing and outputs a presumed superposition noise spectrum using the noise-less voice model memorized by voice model memory, The spectrum operation process which inputs a noise superposition sound signal, performs a analysis of a spectrum, and outputs noise superposition voice spectrum time series, The noise spectrum removal process of removing the spectrum component of a noise using the presumed superposition noise spectrum which is the output of said superposition noise presumption process to the noise superposition voice spectrum time series which is the output of said spectrum operation process, The feature-vector operation process which calculates a feature vector from the noise rejection voice spectrum time series which is the output of said noise spectrum removal process, and outputs noise rejection voice feature-vector time series, As opposed to the noise rejection voice feature-vector time series which is the output of said feature-vector operation process The bottom speech recognition approach of the noise characterized by consisting of collating processes which output the voice model for collating with which collating with the voice model for collating memorized by reference model memory is performed, and likelihood becomes high most as a recognition result.

[Claim 10] Said superposition noise presumption process considers as an input the noise superposition voice feature-vector time series which is the output of a sonagraphy process. The presumed SN ratio operation process of calculating the presumed SN ratio of each noise superposition voice feature vector using the noise model memorized by noise model memory and the noise-less voice model memorized by voice model memory, According to the presumed SN ratio which is the output of said presumed SN ratio operation process, composition with the noise-less voice model memorized by the noise model memorized by noise model memory and voice model memory is performed. The noise superposition voice model generation process which generates a noise superposition voice model, and the noise superposition voice model which is the output of said noise superposition voice model generation process, The likelihood operation process which performs collating with the noise superposition voice feature vector set as the object of a presumed SN ratio operation in said presumed SN ratio operation process, and is outputted as collating data in quest of likelihood, The weight data operation process of calculating the weight data to the combination of three persons of each noise superposition voice feature vector, a noise model, and a noise-less voice model using the collating data outputted from said likelihood operation process, The weight data outputted from said weight data operation process, and the presumed SN ratio which is the output of said presumed SN ratio operation process, The bottom speech recognition approach of the noise according to claim 9 characterized by consisting of noise spectrum generation processes which generate a presumed superposition noise spectrum using noise superposition voice feature-vector time series and a noise model.

[Claim 11] It is the bottom speech recognition approach of the noise according to claim 10 characterized by for said weight data operation process choosing what has the highest likelihood from collating data while said likelihood operation process calculates and outputs collating data for every combination of a noise model and a voice model, and calculating weight data.

[Claim 12] Said likelihood operation process is the bottom speech recognition approach of the noise according to claim 10 characterized by for said weight data operation process choosing in order two or more things which have high likelihood from collating data while calculating and outputting collating data for every combination of a noise model and a voice model, and calculating weight data.

[Claim 13] While said likelihood operation process calculates and outputs collating data for every combination of a noise model and a voice model, said weight data operation process The bottom speech recognition approach of the noise according to claim 12 which chooses in order two or more things which have high likelihood from collating data, carries out weighting addition of the likelihood obtained with the same noise model, and is characterized by calculating weight data in quest of a noise model with the largest value as a result of this weighting addition.

[Translation done.]